Chapter 1 What is Statistics?

While some of these exercises are simple applications of ideas in the chapter, most are intended to stimulate thinking about issues under study and study design issues. In our own teaching we want our students to automatically begin to think about these types of things when presented with information. These exercises are a first step towards that goal. We are not after right or wrong answers. Many of these exercises work well for stimulating group discussion and there are often students in a class with some knowledge of the subject area who can contribute more informed and perceptive answers than the teachers. In the following we simply give a few sketch ideas.

Exercises for Section 1.1

1. We could take all members of the population in the country at the time who were entitled to vote in national elections. In New Zealand this would exclude the young, the illegal immigrants, those people in prisons and people legally committed to mental hospitals. It would include any other legal residents in New Zealand whether or not they were citizens, and citizens living overseas. You might want to be more, or less, restrictive. In practice, one would probably sample from something like the electoral rolls – that subset of people who fit the eligibility criteria for voting and who have registered to do so.

2. −4. These are entirely for you to draw on your own experience. Similar stories are given in the chapter.

5. The survey was self-selective and therefore not necessarily representative of N.Z. women. The reporter compared the statement"92% of incest cases reported in the survey are European" with "81% of the population is European" and claimed unusually high abuse. The reporter should have compared the statement "92% of incest cases reported in the survey are European" with the fact that "91% of those responding to the survey were European." Far from the figure of "92% of incest cases" being an abnormally high proportion, it is almost exactly what you would expect if incest and ethnicity were unrelated.

Exercises for Section 1.2

1. We would choose 10 rats at random to form the reward group. The remaining 10 form the punishment group. We might expect the first 10 rats caught to tend to be the slowest (and perhaps the most stupid!).

2. Product characteristics tend to vary over time, so it is possible that the two lengths chosen may differ in important ways from the other two. Another method would be to randomly choose two of the four rods available from each length for testing process 1 and the reminder for process 2. This would protect us from the possibility of systematic changes along the length of a rod. Alternatively we can use a systematic method: Choose rods 1 and 3 from each of the lengths for process 1, and rods 2 and 4 for process 2.

3. Important things are random selection and double blinding (cf. school milk story). The definition and measurement of intelligence is a thorny problem that has exercised psychologists for years. If one suspected that there would be an effect on a particular type of intelligence, that is the type that should be measured. Otherwise, you would probably just choose a standard broad-spectrum psychological test appropriate for the age group.

Exercises for Section 1.3

Note: In each of these these items there are other possible explanations besides the "obvious" one. Coming up with alternative possibilities is an important part of statistical thinking. Our answers are not the "right" answers. We have no particular expertise in the areas under discussion and just report a few ideas that occurred to us.

1. The "obvious" inference is that with the coming of electricity and the facilities it brings, there are more distractions, and birth rates fall. The observed differences in rates, however, may be related to economic issues, with people not yet having electricity tending to be poorer and less educated. There are probably rural/urban differences. Rural areas tend to be slower to get electricity, and there are practical reasons for rural people to want larger families.

2. The observed difference between the groups may be due to something that is associated with smoking but not caused by smoking. Perhaps people who choose to smoke tend not to work as hard on average, so that it is this and not the smoking that gives them the lower marks. Perhaps their average IQ is lower.

3. Not necessarily. Even long trips have a close-to-home portion. Most of our travelling experience is close to home, so we might expect more accidents close to home.

4. It could be the other way around. Maybe prolonged exposure to alcohol causes these chemical changes.

5. A host of factors may be at work here. Few would doubt that more men are involved in crime than women, so it is not surprising that more end up in jail. It is possible that mothers with dependent children are less likely to be sentenced to jail. The types of crime that men and women tend to commit may also differ.

6. A randomized experiment is not feasible. Subject choice affects the careers you can go into, for example. We cannot direct some randomly chosen people to head off in one direction and others in another. If the gender differences in mathematics scores disappear when we look just at people who also take physics, this would tend to support the "reinforcement" idea. A problem with this is that there may be gender differences in subject choice. Perhaps girls taking physics are better than boys on average because they are less likely to take the subject unless they know they are good at it. (This is easily investigated – how?) Any effect like this would bias the male-female comparison of mathematics scores confined to physics students.

Review Exercises 1

1. **(a)** One definition would be anyone not in full-time paid employment who wants a job and is actively seeking one. This is problematic in that it excludes the long-term unemployed who are so discouraged that they have given up on finding a job.

(c) The definition of what it means to be unemployed differs between countries as does the way the data are collected, for example, surveys (with many differences in survey design) versus official registers of unemployed people.

3. The experiences giving rise to anecdotal evidence are very small samples that may also have been resulted from a very biased "sampling" mechanism.

5. (i) Big unrepresentative samples are still unrepresentative. (ii) This is fine so long as you see it as raising issues rather than representing popular opinion or experience. (iii) Pointing to other instances of unrepresentative research does not make your research any better.

7. **(a)** There will be people with Celtic names but very little Celtic blood (e.g. a Celtic surname could have come from a sole Celtic male ancestor six generations ago). Similarly, there would be people who were predominently Celtic but have non-Celtic names.

(b) Including some non-Celts in with the Celtic group and some Celts in the non-Celtic group, would make the observed differences between the groups *smaller* than they should be.

(c) Psychiatric patients of the type that attended the particular hospital at that time.

(d) Anything that was special about the types of patients admitted and the conditions that they were exposed to.

9. Fatalities per kilometer per car tell us how safe the road is for us to travel on. It tells us nothing, however, about the number of lives that we are likely to be able to save per dollar spent on highway improvement. Short stretches of road with many fatalities, like the bridge, may be good to target in this regard. Fatalities per kilometer of highway may give a reasonable approximate measure.

11. Elevation, shading, steepness, and soil conditions are all factors to be considered. Laying out a grid of cells of equal area on a map of the area and randomly choosing cells to receive each of the three treatments is a reasonable way to proceed if practically feasible.

13. Different places use different definitions. Reporting rates may also differ for a variety of reasons. We have to be a little suspicious that people with a point to prove may be very selective in what they report and how they report it!

15. The paint surfaces that last longest at high temperatures may be different from the surfaces that last longest under normal operating conditions.

Chapter 2 Tools for Exploring Univariate Data

Exercises for Section 2.1

2. Income (continuous), religion (categorical), ethnicity (categorical), distance (continuous), transport (categorical), household number (discrete), smoking (ordinal), cigarettes (discrete), political party (categorical), and percentage income (continuous, though if the data are sufficiently rounded to a few significant figures, it could be treated as discrete).

3. Patient 69 did not continue to smoke, had surgery for symptoms within one year, and was alive at last follow-up.

4. Patient 201 was aged 42 at admission, had DIAVOL = 329, was not taking beta blockers, and had a cholesterol score of 39.

5. Patient 203 was aged 54 at admission, had occlusion score 0, was not taking beta blockers, and had no cholesterol score recorded.

6. Only five patients continued to smoke. Only two had surgery for symptoms after five years. Thirty two were still alive at last follow-up.

Exercises for Section 2.2

Table 2.1 : Simplified World Gold Production Table
(in millions of troy ounces)

Year	World prod.	S. Afr.	U.S.	USSR	Aust.	Can.	China	Col.	Ghana	Phlp.	Mexico	Zaire
1972	45	29	1.4	–	0.8	2.1	–	0.2	0.7	0.6	0.1	0.1
1974	40	24	1.1	–	0.5	1.7	–	0.3	0.6	0.5	0.1	0.1
1975	38	23	1.1	–	0.5	1.7	–	0.3	0.5	0.5	0.1	0.1
1976	39	23	1.0	–	0.5	1.7	–	0.3	0.5	0.5	0.2	0.1
1977	39	23	1.1	–	0.6	1.7	–	0.3	0.5	0.6	0.2	0.1
1978	39	23	1.0	–	0.7	1.7	–	0.3	0.4	0.6	0.2	0.1
1979	39	23	1.0	–	0.6	1.6	–	0.3	0.4	0.5	0.2	0.1
1980	39	22	1.0	8.4	0.6	1.6	–	0.5	0.4	0.8	0.2	0.0
1982	43	21	1.5	8.6	0.9	2.1	1.8	0.5	0.3	0.8	0.2	0.1
1983	45	22	2.0	8.6	1.0	2.4	1.9	0.4	0.3	0.8	0.2	0.2
1984	46	22	2.1	8.7	1.3	2.7	1.9	0.8	0.3	0.8	0.3	0.1
1985	49	22	2.4	8.7	1.9	2.8	2.0	1.1	0.3	1.1	0.3	0.1
1986	52	21	3.7	8.9	2.4	3.4	2.1	1.3	0.3	1.3	0.3	0.2
1987	53	19	4.9	8.9	3.6	3.7	2.3	0.9	0.3	1.1	0.3	0.1
1988	58	20	6.5	9.0	4.9	4.1	2.5	0.9	0.4	1.1	0.3	0.1
1989	64	20	8.5	9.2	6.5	5.1	2.6	0.9	0.4	1.1	0.3	0.1
1990	68	19	9.3	9.7	7.9	5.4	3.2	0.9	0.5	0.8	0.3	0.1
1991	67	19	9.3	7.7	7.5	5.6	3.9	1.0	0.8	0.8	0.3	0.1

1. **(a)** We have rounded severely and ordered the table by importance as a gold producer in the most recent year (1991) to get Table 2.1. We would have preferred to have years as columns and countries as rows as we are most interested in comparing countries. However, this would have forced us to use two pages to represent so many columns, or to suppress some less important producers. We have left the world figures in column 1 so it is easier to compare the figures for the big producers with the world total. South Africa has been by far the biggest producer, but its production figures have been decreasing a little over time. Trends include a gradual increase in world production since 1980. The major increases in production have been in the United States, Australia, Canada, and China. Figures for the USSR start in 1980, but given the large production occurring then, we imagine that they must have been producing substantial amounts before this (similarly for China).

(b) (Using the original table) 1972: 0.65 (65%); 1980: 0.55 (55%); 1986: 0.4 (40%).

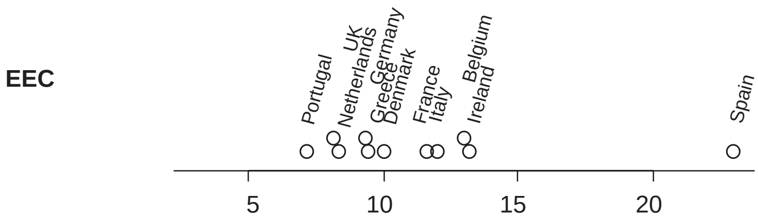
(c) (Using the original table) 1972: 3.2%; 1980: 2.5%; 1986: 7.3%; 1991: 13.9%.

(d) 1970s: Canada; 1980s: USSR; 1990s: U.S.

2. For the reader.

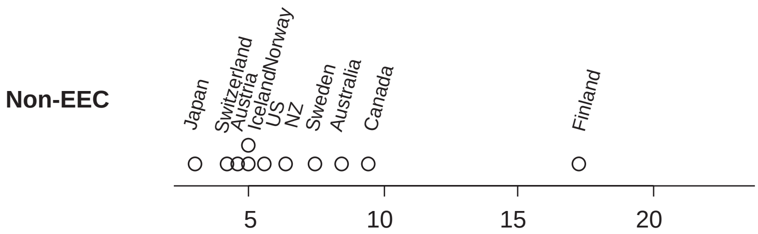
Exercises on Section 2.3.1

1.



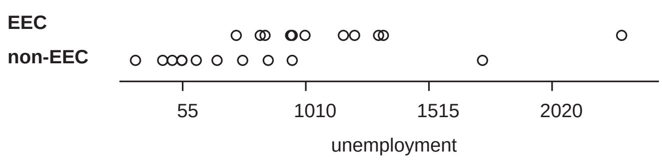
Spain stands out as a high-unemployment outlier.

2.



Finland stands out as a high-unemployment outlier.

3. Plotting against the same scale allows us to compare unemployment patterns of members and non-members. We see that EEC countries had higher unemployment on average than non-EEC members. The spreads are similar.



Exercises for Section 2.3.2

1. **(a)** (i) 1.54. (ii) 0.67. **(b)** (i) 154. (ii) 67. **(c)** (i) 0.00154. (ii) 0.00067.

2. **(a)** $16|3 = 1.63$ cm. **(b)** Round to two significant figures, i.e., $5|4 = 540$ m. **(c)** Round to one decimal place, i.e., $16|3 = 16.3$ m. **(d)** $166|3 = 1663$ kg.

3. 5.4, 9.8, 10.1, 10.1, 10.3, ... , 25.6, 26.8.

4. `Units 0 | 9 = 90`

```
0 | 9
1 | 01122233333334444444
1 | 5555566666667779
2 | 0224
2 | 6
3 | 13
```

(a)

```
0 | 9
1 | 011
1 | 2223333333
1 | 444444455555
1 | 66666666777
1 | 9
2 | 0
2 | 22
2 | 4
2 | 6
2 |
3 | 1
3 | 3
```

(b)

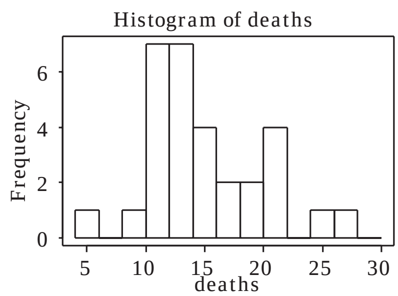
We see a positively skewed shape (long tail toward large values) with a mode at about 145.

6. A stem-and-leaf for position 7 is given, together with other plots, in Fig. 3.2.5 in the main text. The body of the plot is bimodal, and there is a group of three much smaller observations.

Exercises for Section 2.3.3

1.

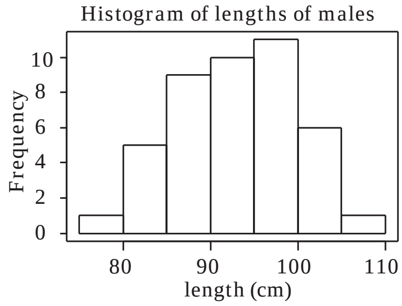
Interval	Freq.
4 6$^-$	1
6 8$^-$	0
8 10$^-$	1
10 12$^-$	7
12 14$^-$	7
14 16$^-$	4
16 18$^-$	2
18 20$^-$	2
20 22$^-$	4
22 24$^-$	0
24 26$^-$	1
26 28$^-$	1



The endpoints of the intervals are slightly different from the stem-and-leaf in Fig. 2.3.7(b), so the frequencies are not identical. The shapes are very similar, however.

2.

Interval		Freq.
75	80$^-$	1
80	85$^-$	5
85	90$^-$	9
90	95$^-$	10
95	100$^-$	11
100	105$^-$	6
105	110$^-$	1



To compare the histograms we would place one histogram above the other and use the same scale. The histogram for the males is shifted to the right, indicating their greater lengths. The one for the females is more peaked.

3. 20/40, or 50%.

Exercises for Section 2.3.4

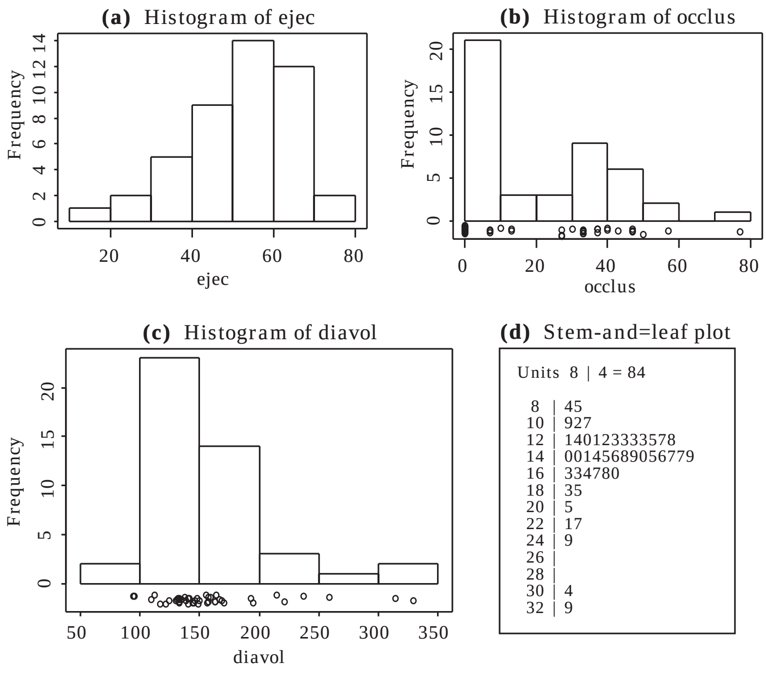


Figure 1 Plots for Exercises for Section 2.3.4

1. See Fig. 1(a). The shape is negatively skewed (long tail to the left).

2. See Fig. 1(b). The shape is bimodal with an outlier.

3. The stem-and-leaf plot given in Fig. 1(d) is very similar in shape to the histogram given in Fig. 1(c). The overall shape is positively skewed (long tail to the right) with possibly two outliers.

Note: In Figs 1(b,c) we have included the dot plot (not part of the histogram) just to reinforce how the data points are represented in the histogram and how the histogram represents relative density. Data points falling on a boundary are included in the class interval above them.]

Exercises on Section 2.4.1

1. **(a)** 7.5 [$(n+1)/2 = 3.5$, so midway between 3rd and 4th largest observations].

(b) 4 [$(n+1)/2 = 4$, so 4th largest observation)].

(c) 3.0 [$(n+1)/2 = 4.5$, so midway between 4th and 5th largest observations].

2. For the female coyotes, Mean = 89.24 and Median = 89.75 (original data) or 90 (stem-and leaf). Data from the stem-and-leaf plot are rounded. No.

3. For the male coyotes, Mean = 92.06 and Median = 92 (either original data or rounded data from a stem-and-leaf). From the samples, the average length of male coyotes is greater than that for the female coyotes.

4. Mean = 48.43. The stem-and-leaf plot for the ages of those with OUTCOME=0 and SURG=0 follows. There are 23 observations.

```
Units 4 | 5 = 45

  3 | 689
  4 | 23
  4 | 556777899
  5 | 11144
  5 | 7889
```

Taking the (23+1)/2 = 12th largest observation gives us Median = 48.

Exercises for Section 2.4.2

1. **(a)** (Min, Q_1, Med, Q_3, Max) = (1, 2, 8, 11.5, 14).

(b) (Min, Q_1, Med, Q_3, Max) = (1, 5, 10, 13, 20).

2. (6.8, 8.45, 8.75, 9.2, 10.4).

3. (36, 45, 48, 54, 59).

4. You would be moderately tall and thin.

Note: The quartiles given here apply the definitions given in the book exactly. Different computer packages use slightly different definitions and will give slightly different answers in some cases. The important thing about quartiles is the idea of a quartile, not fine differences in definitions.

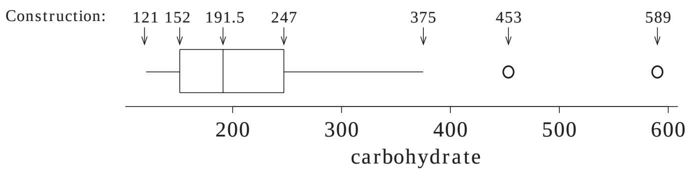
Exercises for Section 2.4.3

1. There are eight of them with diastolic volumes 133, 138, 193, 149, 124, 167, 156, 195. These data have $\overline{x} = 156.875$, $s_X = 26.541$.

2. **(a)** $\overline{x} = 92.056$, $s_X = 6.696$. They are both larger for males. The male data set is slightly more variable. Although on *average* males are longer than females, when it comes to individual comparisons we could not be sure which one would be longer.
   (b) (1 sd) $29/43 = 0.67$ of 67%, (2 sd) $42/43 = 0.98$ or 98%.
   (c) Range $= 105 - 78 = 27$, IQR $= 96 - 87 = 9$. sd/IQR $= 0.744$, or almost exactly 75%.

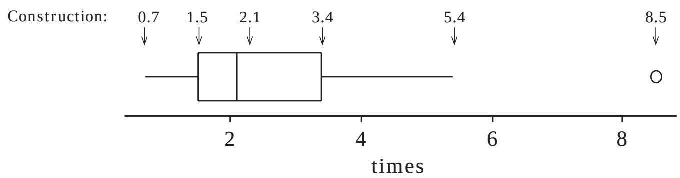
Exercises for Section 2.4.4

1.



The data are positively skewed (long tail to the right) with two outside values (453, 589).

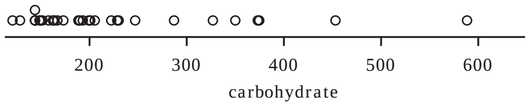
2.



The data are positively skewed with one outside value (8.5).

Notes:

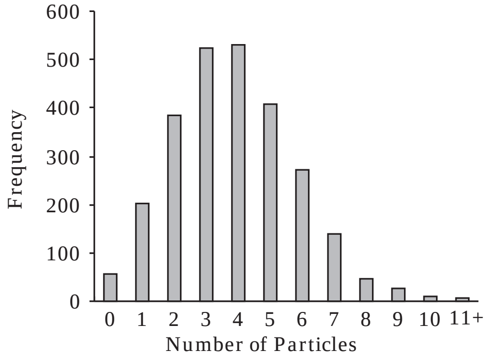
1. In each box plot in Fig. 2.5, the numbers used in the construction of the box plot have been represented just above the plot. The plots have been constructed using the definitions (including the definitions of quartiles) exactly as given in the text. Different computer packages use slightly different definitions so you will often see minor differences in the plots produced by different packages.
2. The rules defining what values are to be plotted as outside values were worked out assuming a Normal distribution (which has a symmetric bell shape). With skewed data, we often see some scattered outside values on the long-tailed side of the distribution. Very often, as here, they are not outliers in the sense that they are so discrepant that we expect they are wrong. They are just part of the skewed shape as the following dot plot of the data in question 1 shows.



Exercises for Section 2.5.1

1. **(a)** Mean $= \frac{0\times57+1\times203+2\times383+...+11\times6}{2608}$ $= 3.870$.

(b) The bar graph as given:



(c) The table is now:

No. of Particles	Freq	Prop.	Percent	Cum. Percent
0	57	0.022	2.2	2.2
1	203	0.078	7.8	10.0
2	383	0.147	14.7	24.7
3	525	0.201	20.1	44.8
4	532	0.204	20.4	65.2
5	408	0.156	15.6	80.8
6	273	0.105	10.5	91.3
7	139	0.053	5.3	96.6
8	45	0.017	1.7	98.3
9	27	0.010	1.0	99.3
10	10	0.004	0.4	99.7
11+	6	0.002	0.2	99.9
Total	2608	0.999	99.9	

(d) 20.4.% **(e)** 65.2.% **(f)** 6.

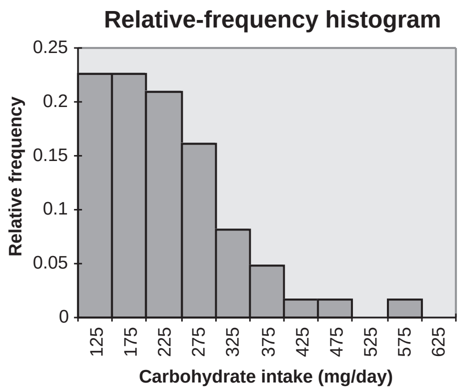
2. They don't make much sense. For example, how would we define a median? Of the observations, 44.8% are smaller than 4, 20.4% are equal to 4, and 34.7% are bigger than 4.

Exercises for Section 2.5.2

1. Mean$=\frac{125\times14\ +\ 175\times14\ +\ 225\times13\ +\ \ldots\ +\ 575\times1}{62} = 227.42$ and sd $= 94.27$ (computer). These answers are fairly close to the answers for the ungrouped data.

2.

From	To	Midpt	Freq.	Rel. Freq.
100	150	125	14	0.2258
150	200	175	14	0.2258
200	250	225	13	0.2097
250	300	275	10	0.1613
300	350	325	5	0.0806
350	400	375	3	0.0484
400	450	425	1	0.0161
450	500	475	1	0.0161
500	550	525	0	0.0000
550	600	575	1	0.0161
		Total	62	1



The shape looks like the right half of a bell. **(a)** 51/62. **(b)** 9/62. **(c)** 2/62.

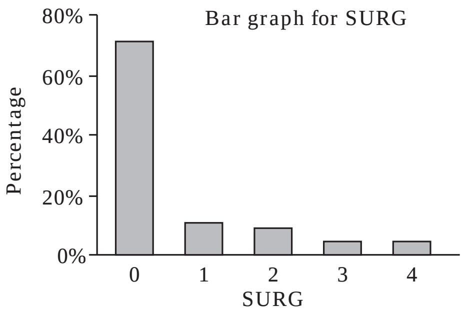
3. The height of the rectangle above the interval 400–600 would be $3/(4 \times 62)$.

Exercises for Section 2.6

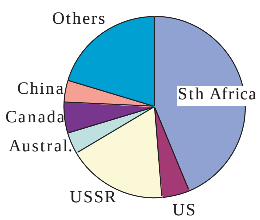
1. The table and graph follow:

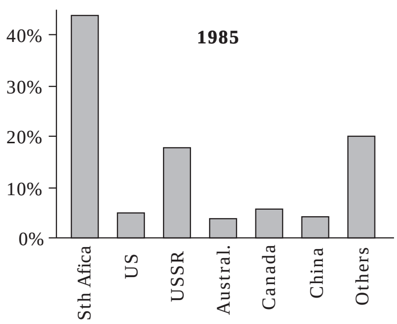
Frequency Table for SURG

SURG	Freq	Percent	Cum. %
0	32	71.111	71.11
1	5	11.111	82.22
2	4	8.889	91.11
3	2	4.444	95.56
4	2	4.444	100.00
	45	100	



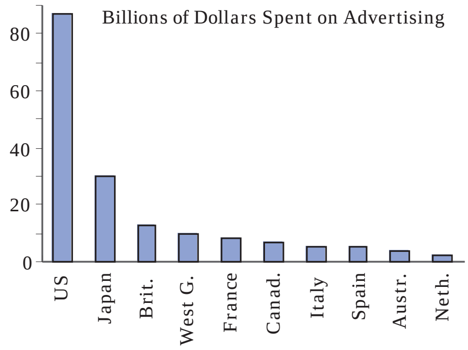
2.



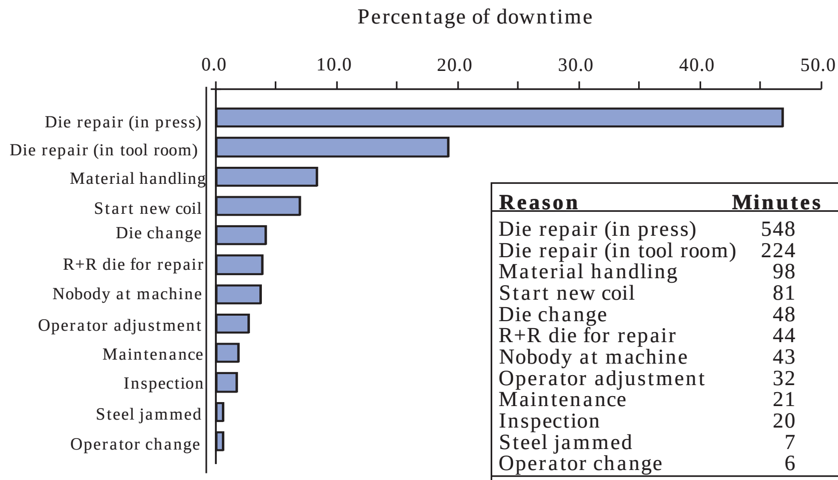


The biggest changes are the reduction in the South African share, and the growth in the US and Australian shares between 1985 and 1991.

3. We use a "decreasing" bar graph.



4. We have used percentage of total downtime as our measure, and ordered by descending percentage. We decided to turn this graph on its side because this is good for displaying long labels. The table containing our working is placed awkwardly to save space. We see that the first two reasons (both die repairs) account for 70% of all downtime. It is here we look first.



Reason	**Minutes**	**Percent**
Die repair (in press)	548	46.8
Die repair (in tool room)	224	19.1
Material handling	98	8.4
Start new coil	81	6.9
Die change	48	4.1
R+R die for repair	44	3.8
Nobody at machine	43	3.7
Operator adjustment	32	2.7
Maintenance	21	1.8
Inspection	20	1.7
Steel jammed	7	0.6
Operator change	6	0.5
Total	1172	

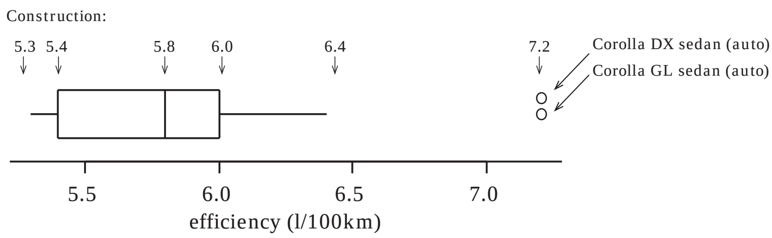
Review Exercises 2

1. **(a)** We have presented two plots, the second being a stretched out version of the first. We see a slightly positively skewed shape with two outlying large values.

```
Units 5 | 3 = 5.3 l/(100ml)
                                  5 | 3
    5 | 344444                    5 | 44444
    5 | 6677899999                5 | 6677
    6 | 124                       5 | 899999
    6 |                           6 | 1
    7 | 22                        6 | 2
                                  6 | 4
                                  6 |
                                  6 |
                                  7 |
                                  7 | 22
```

(b) Med = 5.8, $IQR = Q_3 - Q_1 = 6.0 - 5.4 = 0.6$.

(c)



Both show a central group with two large outliers.

(d) The outliers are the only two cars with automatic transmissions.

3. **(a)** *Home*: mean = 20.227, sd = 6.510; *away*: mean = 14.364, sd = 5.104. There were nearly six more goals scored on average for away games. The spread (sd) is also somewhat greater at home.

(b) Use back-to-back stem-and-leaf plots.

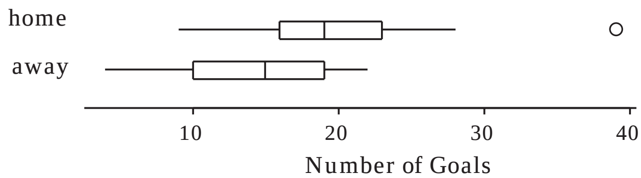
```
Units  1 | 9 = 19 goals
   HOME         AWAY

          | 0 | 4
         9| 0 | 688
       432| 1 | 002334
  99988865| 1 | 555688999
    433220| 2 | 112
       877| 2 |
          | 3 |
         9| 3 |
```

These give the gives same basic impression.

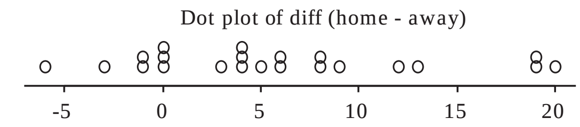
(c) *Home*: (Min, Q_1, Med, Q_3, Max) = (9, 16, 19, 23, 39).

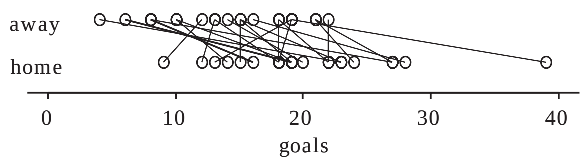
(d)

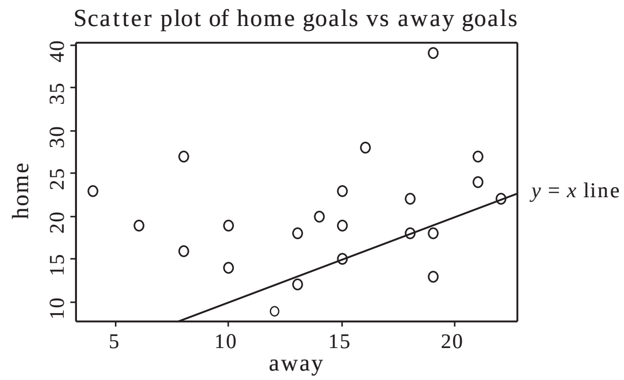


We see the shift toward higher values for home games, the long lower whisker of the "away" box coming from the negative skew in the "away" stem-and-leaf plot. The unusual home-games score of 39 shows up in both the box plot and stem-and-leaf.

(e) We need ways that let us know which pairs of home and away points belong together. We mention three ways. First, we can use a dot plot of the differences. This shows that the home−away differences are almost always positive (i.e., most teams got more goals at home than away). It also shows the variability in the differences. We see 3 teams (Blackburn Rovers, Manchester City & Everton) with unusually large home−away differences. Second, we can use a linked dot plot. Here points that belong together are linked by lines. We can see that the home score is almost always bigger than the away score. This plot also displays the variability in home scores and in away scores. The variability in the differences is harder to see. Third, we can use a scatter plot (discussed in Chapter 3) to show a tendency for teams who get more away goals to also get more home goals. A point would lie on the $y = x$ line if the number of home and away goals were identical. Most points lie considerably above the line (more home goals than away).





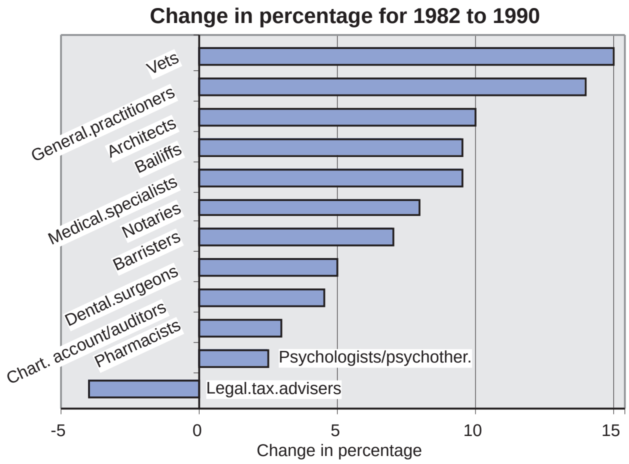


5. **(a)** 31%.

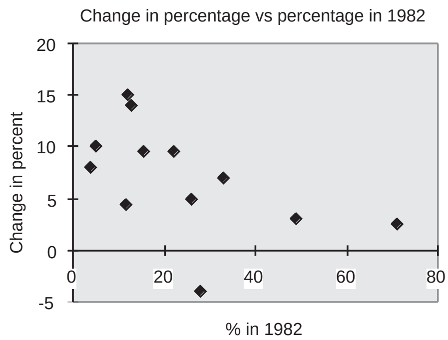
(b) We have constructed the table, ordered by decreasing *change in percentage* (1992 value − 1982 value), and constructed the bar graph. Both are given below. The biggest increases in female participation were for vets and general practitioners (family doctors). There was actually a decrease for legal tax advisers.

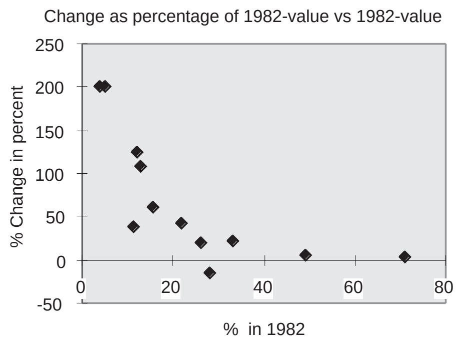
Women in the professions

Profession	%.1982	%.1990	change
Vets	12	27	15
General.practitioners	13	27	14
Architects	5	15	10
Medical.specialists	22	31.5	9.5
Bailiffs	15.5	25	9.5
Notaries	4	12	8
Barristers	33	40	7
Dental.surgeons	26	31	5
Charted.accountants.auditors	11.5	16	4.5
Pharmacists	49	52	3
Psychologists.psychotherapists	71	73.5	2.5
Legal.tax.advisers	28	24	-4



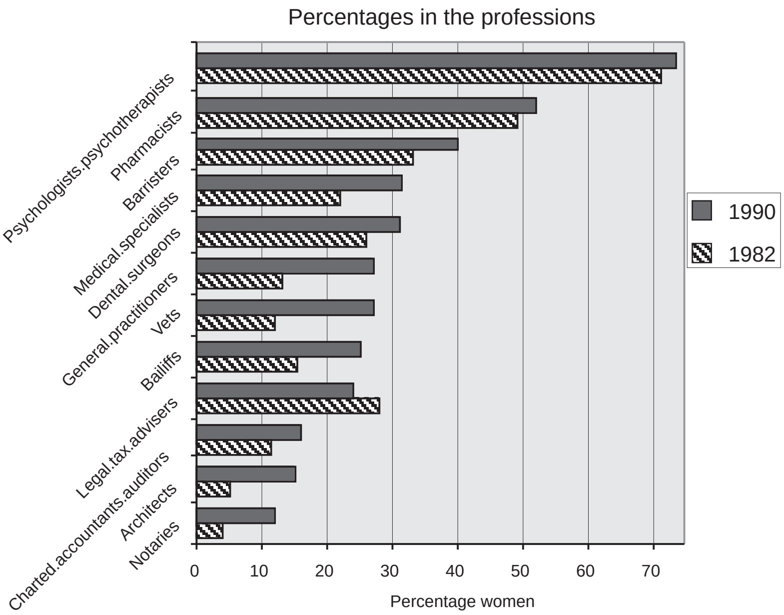
***(c)**





We have given two plots. It depends on what is meant by "gain", as it can be measured by the actual change or the percentage change. Either of these can be plotted against the 1982 percentage using a scatter plot (Chapter 3). Both plots support the conjecture, though the latter does so more strongly.

(d) Below we give a side-by-side bar graph with the two years together for each profession (a standard form of chart from Excel). The plot has been ordered so that the biggest professions with biggest female participation are at the top. We see that the psychologist-physiotherapist category is over 70% female. We also see the biggest changes occurring in the professions in the bottom two thirds of the plot.



7. **(a)** The design seems reasonable. We could also have used BABA. To reduce the amount of change, ABBA or BAAB could have been used.

(b) We see that 738 were caught with mesh size= 35, and 787 were caught with mesh size = 87. Slightly more were caught with the larger mesh, whereas we would expect more to be caught with the smaller mesh (fewer escaping through the mesh). The most obvious explanation is that the boat just happened to come upon fewer fish when trawling the larger-mesh cod end.

(c) Using the rule for estimating means and standard deviations from grouped data, we have: (*35 mm mesh*) mean = 33.427, sd = 3.418; (*87 mm mesh*) mean = 34.549, sd = 3.154. This tells us that the fish caught by the larger-mesh net are slightly longer on average and slightly less variable in length. This is what we should expect because the larger mesh is failing to hold some of the smaller fish. The 1- and 2-sigma rules would lead us to expect that, for each mesh size, about 68% of the fish caught would have a length within $mean \pm sd$, and about 95% would fall within $mean \pm 2sd$. (In fact, these approximations work quite well with these data).

(d) See Fig. 2(d). Relative frequency histograms let us compare the distributions of fish caught (e.g., the proportions caught within a given length range) by the two types of net in a way that is valid, even if the numbers swimming into the net are quite different.

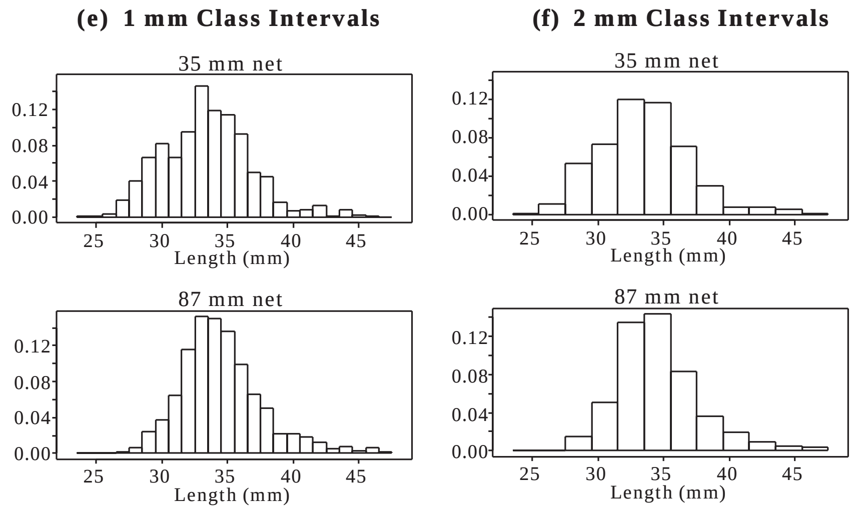
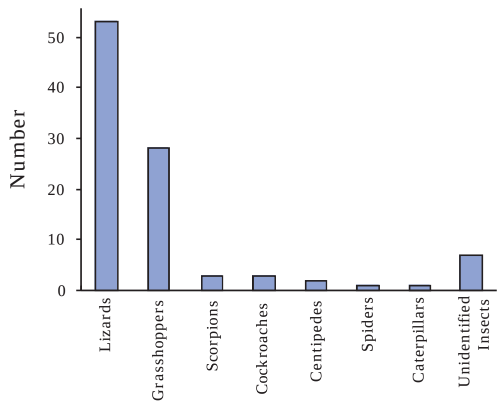


Figure 2 Plots for Question 7(e) and (f)

(e) See Fig. 2(e). The 35 mm mesh caught a higher proportion of smaller fish while the 87 mm mesh caught a higher proportion of larger fish, as expected. The 87 mm mesh would allow more of the smaller fish to escape.

(f) We come to similar conclusions though perhaps it is now easier to see the difference.

9. (a)



Lizards and grasshoppers are very high on the diet list. Between them they account for almost all of the animals eaten.

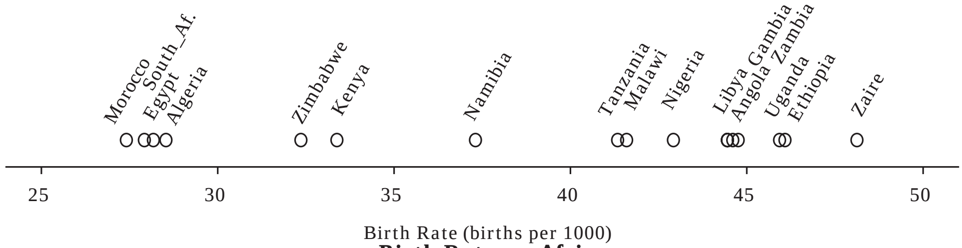
(b) Percentages, as the total numbers of stomachs are different for the two species.

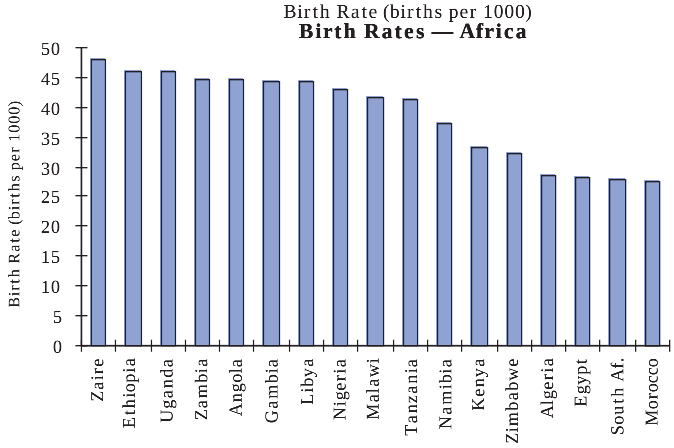
(c) No, because some of the percentages are too small to show on a pie chart.

(d) Depends on the purpose of the study. However, we suggest numbers, as they measure the number of "hits".

(e) Lizards could be more abundant. Preference studies are not easy to carry out. However, the naive approach would be to put *Varanus eremius* in an enclosure with different prey species and see what happens. (However, there can be problems relating to density-dependence which are too technical to discuss.)

11. There is an enormous range in life expectancy, with the highest being Morocco (69.5 years) and the lowest Malawi (36.2 years). The data are fairly well spread out with a hint of small clusters of two or three countries about every five years. These clusters could be studied to see if there are any common features.





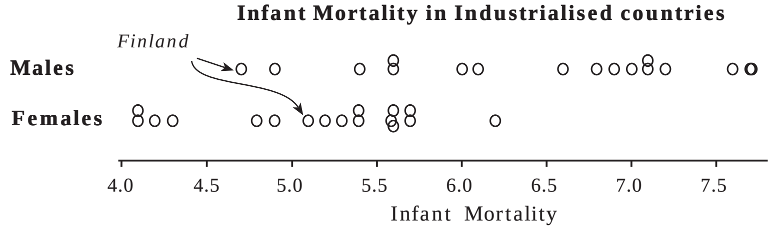
13. **(a)** The female life expectancy is 3 years longer than for males. Zero is special because it is the equality value, namely the female life expectancy is the same as the male). Positive values correspond to a female life expectancy longer than the male.

(b) The female life expectancy is 1.2 times as long as for males (i.e., 20% longer). Unity is special because that is the equality value (female life expectancy same as the male). Values greater (resp. smaller) than 1 correspond to female life expectancies longer (resp. shorter) than male.

15. **(a)** Age structure of the population, availability of birth control methods, infant mortality rate, and so on.

(b) The issues are complex. One reason for the higher death rates might be that the industrialized countries generally have higher life expectancies, which means there are more old people. This could apply to the Central and South American countries as well as the less industrialized Asian countries. However Hong Kong, Japan, and Singapore have lower death rates as well as lower infant mortality rates. The infant mortality rate will affect both the total death rate and the life expectancy, so that these factors are all interrelated.

(c) Finland is different in two respects: It has the lowest male mortality rate, and this rate is less than its female mortality rate.



(d) The mean is not very informative as the data set is small and the numbers quite variable. We have mean = 16.9 and median = 16.3. The median is more informative as it is less affected by the extreme values.

(e) If the emphasis is on global improvement, we should look at the number of people experiencing economic growth, rather than the number of countries.

(f) $Increase = Migrants + Births - Deaths.$